How Serious is the Paradox of Serious Possibility?

Simone Duca

University of Bristol

FEW 2010, Konstanz

Outline of the talk

- Belief change and modalities
- The paradox of serious possibility
- A possible way of defusing the paradox

(This is based on a joint paper with Hannes Leitgeb, which is currently under review.)

Rational agents and revised belief sets

Standard AGM assumes that

- an agent is *rational* iff she obeys certain rationality postulates on her belief sets and her revision operator $*$.
- a *belief set* K is a logically closed set of sentences of an underlying object language $\mathcal{L}$, representing the propositions that a rational agent currently believes.
- provided that A is consistent with K, the *revised* belief set $K^{*}A$ is defined as the deductive closure of $K \cup \{A\}$; otherwise, principles of minimal mutilation are in operation (more on this later).

Introspective agents

Let our object language $\mathcal{L}$ include the following modal operators ($\mathcal{L}$ does not need to be fully closed under them, since we are not directly concerned with iterated modalities):

- $\Box A \Leftrightarrow$ "It is believed that A"
- $\Diamond A \Leftrightarrow$ "It is not believed that $\neg A$"

For any given belief set K, let $Poss(K)$ be the *set of serious possibilities* with respect to K, satisfying:

(A) $\forall A : A \in K \rightarrow \Box A \in Poss(K)$

(B) $\forall A : A \notin K \rightarrow \Diamond\neg A \in Poss(K)$

(C) $Poss(K)$ is the smallest set satisfying conditions (A)-(B).

Introspectively closed belief sets

An *introspective* agent incorporates the set of serious possibilities $Poss(K)$ with respect to her current belief set K in K itself.

Let us then assume that, for any K

(Inc) $Poss(K) \subseteq K$

This assumption yields interesting consequences – as we shall see shortly.

Qualitative postulates for introspective belief revision

(IC) $\exists K, \exists A : \neg(A \in K \vee \neg A \in K)$ (*Incompleteness*)

(P) $\forall K, \forall A : \neg A \notin K \rightarrow K \subseteq K^*A$ (*Preservation*)

(Poss) (a) $\forall K, \forall A : A \in K \rightarrow \Box A \in K$ (*PositiveIntrospection*)

(b) $\forall K, \forall A : A \notin K \rightarrow \Diamond\neg A \in K$ (*NegativeIntrospection*)

(S) $\forall K, \forall A : A \in K^*A$ (*Success*)

(C) $\forall K, \forall A :$ if $K \neq \Lambda$ and $\nvdash \neg A$ then $K^*A \neq \Lambda$ (*Consistency*)

(DC) $\forall K, \forall A :$ if $\Box A \in K$ and $\Diamond\neg A \in K$ then $K = \Lambda$ (*DoxConsistency*)

- K ranges over all possible belief sets. The class of belief sets is assumed to be closed under revision.
- Λ is the "trivial" belief set that is identical to the whole language.
- A ranges over all *factual*, i.e. non-modal *formulae*.

The paradox of serious possibility

e.g. Fuhrmann (1989) and Rott (1989)

Theorem (1)

No belief revision operator $*$ can satisfy all of the postulates (IC), (P), (Poss), (S), (C) and (DC).

Proof of Theorem 1.

(1) $A \notin K$, *from* (IC)
(2) $\neg A \notin K$, *from* (IC)
(3) $\Diamond\neg A \in K$, *from (1), by* (Poss-b)
(4) $\forall B : B \in K \rightarrow B \in K * A$, *by* (P) *and def. of* $\subseteq$
(5) $\Diamond\neg A \in K * A$, *from (3) and (4)*
(6) $A \in K * A$, *by* (S)
(7) $\Box A \in K * A$, *by (6) and* (Poss-a)
(8) $K * A = \Lambda$, *from (5) and (7) with* (DC)
(9) $\nvdash \neg A$, *from (2)*
(10) $K \neq \Lambda$, *from (1)*
(11) $K * A \neq \Lambda$, *from (9), (10) and* (C)
(12) *Contradiction! from (8) and (11)*

□

What's next?

- Prove a probabilistic version of the paradox.
- Show why one should not accept the original formulation of (Poss).
- Make (Poss) syntactically more precise by specifying the context of each doxastic operator, via the introduction of *indices*. (This move may be akin to Lindström (1996)'s in the context of conditionals).
- Finally, show how both the qualitative and probabilistic versions of the paradox are thus defused.

Translation manual

Before moving to the probabilistic counterpart of the paradox, we need a suitable "translation" of certain expressions.

- $\ulcorner A \in K \urcorner \approx \ulcorner P(A) = 1 \urcorner$
- $\ulcorner B \in K^{*}A \urcorner \approx \ulcorner P_A(B) = 1 \urcorner$, where $P_A(B) = P(B|A)$
- $\ulcorner \Box A \urcorner \approx \ulcorner Cr(A) = 1 \urcorner$
- $\ulcorner \Diamond A \urcorner \approx \ulcorner Cr(A) > 0 \urcorner$

Notice that $P(B|A) = \frac{P(A \wedge B)}{P(A)}$, when $P(A) > 0$.

The credence function sign $Cr(\_)$ is the object-linguistic counterpart of belief function signs in the metalanguage.
So, for instance, in order to express the agent's belief that she believes A with degree 1, i.e. that $P(A) = 1$ is the case, we write $P(Cr(A) = 1) = 1$.

Probabilistic postulates

(IC†) $\exists P, \exists A : 0 < P(A) < 1$

(P†) $\forall P, \forall A : P(A) > 0 \rightarrow \forall B[P(B) = 1 \rightarrow P_A(B) = 1]$

(Poss†) (a) $\forall P, \forall A : P(A) = 1 \rightarrow P[Cr(A) = 1] = 1$
(b) $\forall P, \forall A : P(A) < 1 \rightarrow P[Cr(A) < 1] = 1$

(S†) $\forall P, \forall A : P(A) > 0 \rightarrow P_A(A) = 1$

(C†) $\forall P, \forall A : P \neq \Lambda, \nvdash \neg A \rightarrow P_A \neq \Lambda$

(PC†) $\forall P, \forall A : P(Cr(A) = 1) = 1, P(Cr(A) < 1) = 1 \rightarrow P = \Lambda$

- Similarly to the qualitative case, P ranges over all possible belief functions (which are all probability measures, except for one "trivial" function). Also, the class of belief functions is assumed to be closed under conditionalization.
- Λ is the "trivial" belief function defined by $\Lambda(A) = 1$ for all A.
- A ranges over all *factual*, i.e. non-modal *formulae* again.

The paradox rephrased

Theorem (2)

No subjective probability function can satisfy all of the postulates (IC†), (P†), (Poss†), (S†), (C†) and (PC†).

Proof of Theorem 2.

(1) $P(A) < 1$, *from* ($\text{IC}^\dagger$)

(2) $P(A) > 0$, *from* ($\text{IC}^\dagger$)

(3) $P(Cr(A) < 1) = 1$, *from (1) and* ($\text{Poss-b}^\dagger$)

(4) $\forall B[P(B) = 1 \rightarrow P_A(B) = 1]$, *from (2) and* ($\text{P}^\dagger$)

(5) $P_A(Cr(A) < 1) = 1$, *from (3) and (4)*

(6) $P_A(A) = 1$, *from (2) and* ($\text{S}^\dagger$)

(7) $P_A(Cr(A) = 1) = 1$, *by (6) and* ($\text{Poss-a}^\dagger$)

(8) $P_A = \Lambda$, *from (5) and (7) with* ($\text{PC}^\dagger$)

(9) $\nvdash \neg A$, *from (2)*

(10) $P \neq \Lambda$, *from (1)*

(11) $P_A \neq \Lambda$, *from (9), (10) and* ($\text{C}^\dagger$)

(12) *Contradiction*! *from (8) and (11)*

□

Proof of Theorem 2.

(1) $P(A) < 1$, *from* ($\text{IC}^\dagger$)

(2) $P(A) > 0$, *from* ($\text{IC}^\dagger$)

!!!(3) $P(\boxed{\text{Cr(A)}} < 1) = 1$, *from (1) and* ($\text{Poss-b}^\dagger$)

(4) $\forall B[P(B) = 1 \rightarrow P_A(B) = 1]$, *from (2) and* ($\text{P}^\dagger$)

(5) $P_A(Cr(A) < 1) = 1$, *from (3) and (4)*

(6) $P_A(A) = 1$, *from (2) and* ($\text{S}^\dagger$)

!!!(7) $P_A(\boxed{\text{Cr(A)}} = 1) = 1$, *by (6) and* ($\text{Poss-a}^\dagger$)

(8) $P_A = \Lambda$, *from (5) and (7) with* ($\text{PC}^\dagger$)

(9) $\nvdash \neg A$, *from (2)*

(10) $P \neq \Lambda$, *from (1)*

(11) $P_A \neq \Lambda$, *from (9), (10) and* ($\text{C}^\dagger$)

(12) *Contradiction*! *from (8) and (11)*

□

Amended qualitative postulates

($\overline{\text{IC}}$) $\exists K, \exists A : \neg(A \in K \vee \neg A \in K)$

($\overline{\text{P}}$) $\forall K, \forall A : \neg A \notin K \rightarrow K \subseteq K^*A$

($\overline{\text{Poss}}$)
- (a) $\forall K, \forall A : A \in K \rightarrow \Box_K A \in K$
- (b) $\forall K, \forall A : A \notin K \rightarrow \Diamond_K \neg A \in K$

($\overline{\text{S}}$) $\forall K, \forall A : A \in K^*A$

($\overline{\text{C}}$) $\forall K, \forall A : K \neq \Lambda$ and $\nvdash \neg A \rightarrow K^*A \neq \Lambda$

($\overline{\text{DC}}$) $\forall K, \forall A : \Box_K A \in K$ and $\Diamond_K \neg A \in K \rightarrow K = \Lambda$

- We introduced the notation "$\Box_t$", where t is a member of the set of terms $\mathcal{T}$ or names for belief sets, whose inductive definition is given simultaneously with the set of formulae $\mathcal{L}$ (as we do in detail in the paper).
- Here I am a bit more lax, so, for instance, I will use "$\Box_{K^*A}$" in order to refer to the doxastic operator expressing membership in K^*A (and no other belief set).

Pseudo-proof of Theorem 1.

(1) $A \notin K$, *from* ($\overline{\text{IC}}$)

(2) $\neg A \notin K$, *from* ($\overline{\text{IC}}$)

(3) $\Diamond_K \neg A \in K$, *from (1), by* ($\overline{\text{Poss-b}}$)

(4) $\forall B : B \in K \rightarrow B \in K * A$, *by* ($\overline{\text{P}}$) *and def. of* $\subseteq$

(5) $\Diamond_K \neg A \in K * A$, *from (3) and (4)*

(6) $A \in K * A$, *by* ($\overline{\text{S}}$)

(7) $\Box_{K*A} A \in K * A$, *by (6) and* ($\overline{\text{Poss-a}}$)

$\boxtimes$

(8) $K * A = \Lambda$, *from (5) and (7) with* ($\overline{\text{DC}}$)

(9) $\nvdash \neg A$, *from (2)*

(10) $K \neq \Lambda$, *from (1)*

(11) $K * A \neq \Lambda$, *from* (9), (10) and ($\overline{\text{C}}$)

(12) *Contradiction*! *from (8) and (11)*

□

The diagnosis

- $\Box A$ is really short for '*In this world*, it is necessary that A', '*In this belief state*, it is believed that A', and so on.
  In other words, modal operators are context-sensitive.
- In the possible worlds semantics of modal logic, this does not lead to any trouble, for it is not the case that evaluations of $\Box A$ relative to two different contexts ever get mixed up.
  Every evaluation is relative to precisely one possible world, which represents a unique semantic context. Hence, the usual 'elliptical' form works fine.

$\Rightarrow$ However, this "mixing up" of contexts is exactly what is going on in the original derivation of the paradox!
In fact, *within one and the same context, one sign in the object language is meant to stand for two different doxastic states.*
It is this inaccuracy that created the 'illusion' of a paradox.

Amended probabilistic postulates

($\overline{\mathrm{IC}^\dagger}$) $\exists P, \exists A : 0 < P(A) < 1$

($\overline{\mathrm{P}^\dagger}$) $\forall P, \forall A : P(A) > 0 \rightarrow \forall B[P(B) = 1 \rightarrow P_A(B) = 1]$

($\overline{\mathrm{Poss}^\dagger}$)
- (a) $\forall P, \forall A : P(A) = 1 \rightarrow P[Cr_P(A) = 1] = 1$
- (b) $\forall P, \forall A : P(A) < 1 \rightarrow P[Cr_P(A) < 1] = 1$

($\overline{\mathrm{S}^\dagger}$) $\forall P, \forall A : P(A) > 0 \rightarrow P_A(A) = 1$

($\overline{\mathrm{C}^\dagger}$) $\forall P, \forall A : P \neq \Lambda, \nvdash \neg A \rightarrow P_A \neq \Lambda$

($\overline{\mathrm{PC}^\dagger}$) $\forall P, \forall A : P(Cr_P(A) = 1) = 1,\ P(Cr_P(A) < 1) = 1 \rightarrow P = \Lambda$

- Accordingly, the contextual dependence of the Cr function sign can be made explicit.
  Here, Cr_P is an object-linguistic function sign that corresponds to, and denotes, the function P, whilst Cr_{P_A} denotes the function P_A that arises from updating P by A.

Pseudo-proof of Theorem 2.

(1) $P(A) < 1$, *from* ($\overline{\mathrm{IC}^\dagger}$)
(2) $P(A) > 0$, *from* ($\overline{\mathrm{IC}^\dagger}$)
(3) $P(Cr_P(A) < 1) = 1$, *from (1) and* ($\overline{\mathrm{Poss\text{-}b}^\dagger}$)
(4) $\forall B[P(B) = 1 \rightarrow P_A(B) = 1]$, *from (2) and* ($\overline{\mathrm{P}^\dagger}$)
(5) $P_A(Cr_P(A) < 1) = 1$, *from (3) and (4)*
(6) $P_A(A) = 1$, *from (2) and* ($\overline{\mathrm{S}^\dagger}$)
(7) $P_A(Cr_{P_A}(A) = 1) = 1$, *by (6) and* ($\overline{\mathrm{Poss\text{-}a}^\dagger}$)
⊠
(8) $P_A = \Lambda$, *from (5) and (7) with* ($\overline{\mathrm{PC}^\dagger}$)
(9) $\nvdash \neg A$, *from (2)*
(10) $P \neq \Lambda$, *from (1)*
(11) $P_A \neq \Lambda$, *from (9), (10) and* ($\overline{\mathrm{C}^\dagger}$)
(12) *Contradiction! from (8) and (11)*

□

Summary

- We gave a probabilistic version of the paradox of serious possibility.
- We highlighted an important ambiguity in one of the assumptions of the paradox, i.e. (Poss).
- We suggested a possible way of getting rid of the ambiguity and gave two sets (one qualitative and one probabilistic) of amended axioms from which the paradox cannot be derived.
- Furthermore, in our paper, we prove the amended axioms jointly consistent by providing a model, in which the agent is able to "keep track" of her introspective beliefs by a system of indices for belief sets/belief functions.

Thank you